

بر آورد وجه نقد ورودی و خروجی شعب بانک تجارت برای محاسبه وجه نقد مورد نیاز شعبه‌ها با استفاده از تحلیل چندمتغیره خوشه‌بندی بیزی و پیاده‌سازی آن در شبکه‌های عصبی

غزاله باغبانی^۱، فرزاد اسکندری^۲

چکیده: موضوع کفایت وجه نقد در بانک‌ها، یکی از مسائل مهم برای مدیران و به‌خصوص رؤسای هر شعبه به‌شمار می‌رود؛ چراکه کمبود وجه نقد روزانه در صندوق شعبه به عدم پاسخگویی به نیاز مشتری می‌انجامد و از سوی دیگر، مازاد وجه نقد در شعبه موجب افزایش هزینه بابت انتقال آن به خزانه بانک می‌شود. از این رو بانک‌ها همواره درصد تعیین مقدار وجه نقد مورد نیاز خود با توجه به عملیات روزانه هستند. به‌همین منظور در این مقاله، شعب بانک تجارت، با توجه به تنوع بین شعب، با دو روش خوشه‌بندی سلسله‌مراتبی و خوشه‌بندی بر مبنای رویکرد بیزی در خوشه‌های متشابه دسته‌بندی شدند؛ سپس با در نظر گرفتن نتایج خوشه‌بندی، مقدار وجه نقد ورودی و نیز وجه نقد مصرفی از طریق شبکه‌های عصبی برآورد شد تا از این طریق امکان محاسبه وجه نقد لازم برای شعب فراهم شود. نتایج تحقیق نشان می‌دهد، برآورد وجه نقد مصرفی و ورودی شعب بانک با استفاده از شبکه عصبی و لحاظ کردن نتایج خوشه‌بندی شعب با رویکرد بیزی، دارای دقت بیشتری نسبت به نتایج خوشه‌بندی شعب با روش معمول است.

واژه‌های کلیدی: بانکداری، برآورد، خوشه‌بندی، رویکرد بیزی، شبکه عصبی.

۱. دانشجوی دکتری آمار، پردیس دانشگاه علامه طباطبائی، تهران، ایران

۲. دانشیار گروه آمار، دانشکده علوم ریاضی دانشگاه علامه طباطبائی، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۵/۰۴/۱۳

تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۱۰/۲۲

نویسنده مسئول مقاله: فرزاد اسکندری

E-mail: askandari@atu.ac.ir

مقدمه

هدف اصلی در خوشه‌بندی، جای‌دادن مشاهدات در گروه‌ها و دسته‌های متشابه و همگن است؛ به طوری که مشاهدات هر خوشه بیشترین تشابه را با سایر مشاهدات همان خوشه و بیشترین عدم تشابه را با مشاهدات سایر خوشه‌ها خواهند داشت. در این خصوص و به‌ویژه در حوزه بانکداری، شعبه‌ها از لحاظ برخی ویژگی‌ها متفاوت‌اند که می‌توانند بر اساس روش‌های خوشه‌بندی در گروه‌های مشابهی دسته‌بندی شوند و برای برآورد وجه نقد موردنیاز هر شعبه که در خوشه مشابه خود دسته‌بندی شده‌اند، مفید باشند.

این موضوع بدان جهت اهمیت دارد که یکی از دغدغه‌های مهم در تصمیم‌گیری مدیران و به‌خصوص رؤسای شعب، موضوع کفایت وجه نقد است؛ چراکه کمبود وجه نقد روزانه در صندوق شعبه عدم پاسخگویی به نیاز مشتری است و به‌طور مسلم موجب نارضایتی مشتری می‌شود و از سوی دیگر نیز مازاد وجه نقد در شعبه، موجب افزایش هزینه بیمه وجوه و انتقال آن به خزانه بانک، افزایش هزینه خواب پول و از دست دادن فرصت سرمایه‌گذاری برای بانک و به تبع آن عدم کسب درآمد و سود برای بانک است. از این رو بانک‌ها همواره درصدد تعیین وجه نقد موردنیاز خود با توجه به عملیات روزانه هستند.

در حال حاضر، تأمین وجه نقد موردنیاز شعبه در ابتدای همان روز کاری و تنها بر مبنای تجربه رئیس شعبه صورت می‌گیرد که گاهی با عملکرد واقعی آن روز فاصله داشته و می‌تواند به افزایش نقل و انتقال پول از طریق ماشین‌های پول‌رسان مناطق منجر شود که در این صورت هزینه بانک را افزایش می‌دهد، یا در صورت عدم تأمین به‌موقع وجه نقد موردنیاز، موجب نارضایتی مشتریان می‌شود. به همین منظور، لازم است با روش مناسب و کارایی وجه نقد مورد نیاز شعبه‌ها را برای دوره زمانی مدنظر در آینده (برای مثال برای روز کاری بعد)، برآورد کرد تا از مشکلات بیان شده پیشگیری به‌عمل آید.

پیشینه پژوهش

تاکنون از روش‌های مختلفی مانند سری‌های زمانی و شبکه‌های عصبی کلاسیک برای پیش‌بینی وجه نقد مورد نیاز شعبه‌های بانک استفاده شده است؛ به‌طوری که کاردونا و آمایا (۲۰۱۲) طی پژوهشی در بانک کلمبیا، ابتدا با دو روش سری زمانی و شبکه عصبی کلاسیک، وجه نقد شعب بانک را پیش‌بینی کردند؛ سپس به کمک روش برنامه‌ریزی خطی مدل بهتری را انتخاب کرده و به مینیمم نمودن تابع هزینه پرداختند. همچنین در تحقیق دیگری که سندی‌پان (۲۰۱۰) برای شعب بانک ICICI در هند انجام داد، با استفاده از روش سری زمانی به پیش‌بینی

وجه نقد شعبه‌ها پرداخت؛ سپس به کمک روش برنامه‌ریزی خطی به بهینه‌سازی اقدام کرد. پرمچند و والیا (۲۰۰۶) نیز در پیش‌بینی وجه نقد شعبه بانک دولتی هند با دو روش سری زمانی و شبکه عصبی، به برآورد وجه نقد شعبه پرداختند. آنها در تحقیق خود تنها از متغیرهای تقویمی برای پیش‌بینی وجه نقد استفاده کردند.

در بررسی وضعیت تقاضای وجه نقد برای دو شعبه نمونه در بانک کارآفرین، از دو روش سری زمانی و مدل انتشار استفاده شده است که به کمک مدل انتشار و انجام شبیه‌سازی برای توابع توزیع مقدار واریزها و برداشت‌ها، سقف بهینه پول در صندوق شعبه نمونه با کاهش ۲۵ درصدی مواجه شد (نوربخش، حیدری و زواریان، ۱۳۸۹).

بهشتی (۱۳۸۹) در تحقیقی به بررسی پیش‌بینی وجه نقد بانک‌ها با روش شبکه عصبی پرداخت. وی متغیرهایی برای پیش‌بینی، به عنوان ورودی مدل تعیین کرد و فرایند پیش‌بینی را با استفاده از روش‌های اقتصادسنجی، تعریف تابع پیش‌بینی و به کارگیری مدلی با استفاده از شبکه عصبی پرسپترون چند لایه انجام داد. او در تحقیق خود نشان داد که روش شبکه عصبی نسبت به روش‌های دیگر برتری دارد و دارای دقت بیشتری برای پیش‌بینی است. در این تحقیق، متغیرهای تقویمی که بر برداشت وجه نقد مشتریان مؤثرند، در نظر گرفته نشده است.

در این رابطه و به منظور بررسی نتایج به دست آمده از روش‌های متفاوت خوشه‌بندی (شامل خوشه‌بندی بر مبنای مدل و خوشه‌بندی بر مبنای فاصله) برای برآورد وجه نقد مورد نیاز شعب، در این مقاله با استفاده از رویکرد بیزی در خوشه‌بندی بر مبنای مدل، به برآورد وجه نقد ورودی و مصرفی شعب بانک تجارت می‌پردازیم تا از این طریق، وجه نقد مورد نیاز شعب تعیین شود. از جمله مطالعاتی که در خصوص تحلیل خوشه‌بندی بیزی صورت گرفته است، می‌توان به کار فرانزن (۲۰۰۸)، اشاره کرد. او با در نظر گرفتن مدل آمیخته^۱ متناهی برای داده‌ها، با استفاده از روش بیزی به محاسبه توزیع پسین برای پارامترهای هر خوشه از طریق زنجیره مارکوف مونت کارلو (MCMC)^۲ پرداخت. پرتوی نیا (۲۰۰۹) نیز در تحقیقی، رویکرد بیزی در مدل‌های آمیخته را برای داده‌های پیوسته به منظور خوشه‌بندی^۳ و دسته‌بندی^۴ داده‌ها ارائه کرد. در تحقیق او، خوشه‌بندی سلسله‌مراتبی تراکمی^۵ برای ایجاد درخت دندوگرام، بر اساس احتمال پسین خوشه‌ها استفاده شده است. همچنین، بلومزند، تانگ، ژیانگ و گراندلند (۲۰۱۵) در تحقیقی،

-
1. Mixture Model
 2. Markov Chain Monte Carlo
 3. Clustering
 4. Classification
 5. Agglomerative Hierarchical Clustering

استفاده از رویکرد بیزی برای خوشه‌بندی داده‌های آمیخته گسسته و پیوسته را بررسی کردند. آنها با معرفی خوشه‌بندی برمبنای مدل و در نظر گرفتن بردار داده‌های آمیخته از نوع گسسته و پیوسته، از طریق رویکرد بیزی به محاسبه احتمال پسین هر خوشه پرداختند و بهترین خوشه برای داده‌ها را تعیین کردند. از مطالعات دیگر انجام شده در این زمینه، می‌توان به پژوهش برناردو و جیرون (۱۹۸۸) اشاره کرد که آنها نیز با استفاده از رویکرد بیزی در توزیع‌های آمیخته متناهی با در نظر گرفتن تابع توزیع پیشین ناآگاهی‌بخش^۱ برای پارامترهای مدل، به تحلیل خوشه‌بندی پرداختند.

شایان ذکر است که تاکنون در تمام مطالعات انجام شده در حوزه یاد شده با روش‌هایی مانند سری‌های زمانی، محاسبه وجه نقد مورد نیاز تنها برای یک شعبه انجام شده و تعداد و تنوع زیاد شعب مدنظر قرار نگرفته است. از این رو استفاده از رویکرد چند متغیره بیزی در خوشه‌بندی، به‌منظور محاسبه وجه نقد مورد نیاز بانک‌ها که در این مقاله برای چندین شعبه بررسی و مطالعه شده، روشی نوین است که تاکنون در مطالعات داخلی و خارجی، انجام نشده است.

روش‌شناسی پژوهش

از آنجا که بانک تجارت شعبه‌های بسیار زیادی دارد، جامعه آماری این پژوهش، شعبه‌های واقع در تهران مدنظر قرار گرفت. همچنین، با توجه به تنوع شعبه‌ها در هر یک از مناطق تهران و مدیریت‌های شعب مربوطه در بانک، از نمونه‌گیری خوشه‌ای تصادفی برای انتخاب شعبه‌ها استفاده شد که ۱۸ شعبه از طریق نمونه‌گیری برای اجرای پژوهش انتخاب گردید.

متغیرهای استفاده‌شده در خوشه‌بندی براساس نوع تحقیق، متفاوت‌اند. در پژوهش حاضر با توجه به گستردگی و تنوع شعب بانک، بر اساس تعریف متغیرهایی از قبیل درجه شعبه، نوع شعبه از لحاظ سپرده‌ای یا تسهیلاتی، تعداد دستگاه خودپرداز در شعبه، شعبه کشیک / غیر کشیک، به خوشه‌بندی شعب پرداخته شد. شایان ذکر است که متغیرهای یاد شده، براساس مصاحبه با خبرگان بانک استخراج شده‌اند.

گفتنی است که شعبه‌های بانک تجارت بر اساس برخی شاخص‌ها، به درجه‌های ممتاز الف، ممتاز ب، یک، دو، سه، چهار و پنج دسته‌بندی می‌شوند. همچنین، شعبه‌های کشیک تا ساعت ۱۶:۳۰ فعالیت دارند و به تبع آن ممکن است به وجه نقد بیشتری نیاز داشته باشند. نوع شعبه از لحاظ سپرده یا تسهیلاتی نیز براساس میانگین سالانه بدهکاری یا بستانکاری شعب به حساب

1. Non-informative

مرکز تعیین شده است. در ضمن، هرچه تعداد دستگاه‌های خودپرداز در شعبه بیشتر باشد، وجه نقد مصرفی شعبه بیشتری نیاز خواهد بود.

در گام بعد، با مدنظر قراردادن نتایج حاصل از خوشه‌بندی با رویکرد بیزی و شماره خوشه شعب، متغیرهای روزهای هفته (شامل ترتیب روز کاری و چندمین روز هفته)، روزهای پرداخت حقوق / واریز یارانه / واریز سود و نیز روزهای تعطیل و مناسبت‌های رسمی، وجه نقد مصرفی شعبه و نیز وجه نقدی که توسط مشتری وارد شعبه می‌شود، برآورد می‌گردد که از طریق آنها، وجه نقد مورد نیاز شعب پیش‌بینی خواهد شد. شایان ذکر است که دو متغیر وجه نقد مصرفی و ورودی شعبه، به‌عنوان دو متغیر مستقل از هم در نظر گرفته شده‌اند.

داده‌های مربوط به متغیرهای یاد شده در بازه زمانی ۹۳/۲/۱ تا ۹۳/۶/۳۱ برای ۱۸ شعبه بانک که از طریق نمونه‌گیری خوشه‌ای تصادفی تعیین شده‌اند، به‌دست آمده است. خوشه‌بندی داده‌ها بر مبنای رویکرد بیزی، به محاسبات پیشرفته شبیه‌سازی از طریق روش‌های معتبر زنجیر مارکوف مونت کارلو، نیاز دارد که این موضوع و نیز برآورد وجوه نقد مصرفی و آورده مشتری به روش شبکه عصبی، با استفاده از برنامه‌نویسی در نرم‌افزار R انجام شده است.

خوشه‌بندی

خوشه‌بندی یکی از تکنیک‌های پرکاربرد در داده‌کاوی است و برای گروه‌بندی مجموعه‌ای از مشاهدات به کلاس‌های مشابه استفاده می‌شود (هان و کامبر، ۲۰۰۶: ۴۴۳؛ برامبر، ۲۰۰۷: ۲۲۱ و ونگ، ۲۰۰۸). روش‌های خوشه‌بندی به دو دسته روش‌های بر مبنای فاصله و روش‌های بر مبنای مدل دسته‌بندی می‌شوند (پرتوی‌نیا و دیویسن، ۲۰۱۲). از جمله روش‌های پرکاربرد در خوشه‌بندی بر مبنای فاصله که از نوع خوشه‌بندی کلاسیک محسوب می‌شود، می‌توان به روش سلسله‌مراتبی اشاره کرد. در پژوهش حاضر، با توجه به نوع داده‌های مربوط به متغیرهای مرتبط با خوشه‌بندی، برای خوشه‌بندی سلسله‌مراتبی شعب به روش کلاسیک، از ضریب شباهت گوور (۱۹۷۱) به‌عنوان معیار شباهت و تعیین معیار فاصله استفاده شده است. به‌علاوه، در خوشه‌بندی به روش کلاسیک، از معیار فاصله اقلیدسی برای تعیین مشاهدات متشابه و نزدیک به هم استفاده می‌شود و داده‌ها با احتمال قطعی ۰ یا ۱ به یک خوشه خاص منتسب می‌شوند که این موضوع می‌تواند مشاهداتی که نزدیک مرز خوشه‌ها قرار می‌گیرند، را نادیده بگیرد. از این رو استفاده از خوشه‌بندی بر مبنای مدل که بر پایه مدل‌ها و توزیع‌های احتمال است و می‌تواند داده‌های نزدیک مرز خوشه‌ها را با لحاظ مقدار احتمال آنها جهت انتساب به خوشه، مدنظر قرار دهد، مناسب است. از مزیت دیگر روش یاد شده این است که خوشه‌های استخراج شده از دیدگاه

آماري تفسير مي‌شوند (بوويرن و برونٲ، ۲۰۱۲؛ پرتوي نيا، ۲۰۰۹). علاوه بر اين، خوشه‌بندي برمبنای مدل، امکان تعيين تعداد خوشه بهينه را به صورت همزمان فراهم مي‌آورد (بلومز تد و همکاران، ۲۰۱۵). با توجه به اين که در پژوهش حاضر علاوه بر خوشه‌بندي کلاسيک، از خوشه‌بندي بر مبنای مدل با لحاظ رويکرد بيزي استفاده مي‌شود، در ادامه، به توضيح بيشتر آنها پرداخته شده است.

خوشه‌بندي برمبنای مدل

در خوشه‌بندي برمبنای مدل، يک خانواده از مدل‌های آماری برای داده‌ها در نظر گرفته شده و خوشه‌بندي از طريق برازش مدل آميخته^۱ انجام مي‌شود. در اين نوع خوشه‌بندي، فرض بر آن است که داده‌ها از مجموعه‌ی جامعه‌های متناهی جمع‌آوری شده‌اند.

در اين خصوص، فرض کنيد داده‌ها در خوشه^c از مدل پارامتری $f(y_c|\theta_c)$ پيروي کنند. در اين صورت، توزيع کلی $\sum_{c=1}^C p_c f(y_c|\theta_c)$ يک مدل آميخته است و p_c نسبت داده‌های متعلق به خوشه^c خواهد بود. اين مدل، به طور معمول از خانواده‌ی توزيع نرمال يا لاپلاس انتخاب مي‌شود. همچنين، اگر T_c مشاهده در خوشه^c وجود داشته باشد، توزيع داده‌ها مشابه توزيع مشاهدات در خوشه^c است. اگر مشاهدات خوشه^c به صورت دلخواه مرتب شده باشد؛ يعني $f(y_{1c}, \dots, y_{T_c c})$ توزيع تعويض‌پذير^۲ است و بر اساس نظريه کلی جايگزينی^۳، توزيع شرطی $f(y_c|\theta_c)$ و تابع توزيع چگالی پيشين $f(\theta_c)$ وجود دارند، به صورتی که:

$$f(y_{1c}, \dots, y_{T_c c}) = \int \prod_{t=1}^{T_c} f(y_{tc} | \theta_c) f(\theta_c) d\theta_c \quad \text{رابطه (۱)}$$

تعريف ۱: يک دنباله متناهی از متغيرهای تصادفی X_1, \dots, X_n را با تابع توزيع احتمال توأم p تعويض‌پذير گويند، اگر برای هر جايگشت اندیس π داشته باشيم:

$$P(X_1, X_2, \dots, X_n) = P(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}) \quad \text{رابطه (۲)}$$

همچنين، هر دنباله نامتناهی‌ای تعويض‌پذير است که هر دنباله متناهی از آن تعويض‌پذير باشد.

1. Mixture Model
2. Exchangeable Distribution
3. General Representation Theorem

تعریف ۲ (نظریه کلی جایگزینی): اگر X_1, X_2, \dots دنباله تعویض پذیر از متغیرهای با تابع احتمال P باشند، تابع توزیع Q بر F (مجموعه تمام توابع توزیع بر \mathcal{R}) وجود خواهد داشت، به طوری که توزیع توأم (X_1, \dots, X_n) به صورت زیر است.

$$p(X_1, \dots, X_n) = \int \prod_{i=1}^n P(X_i | F) dQ(F) \quad \text{رابطه ۳}$$

که F تابع توزیع نامعین است؛ $Q(F) = \lim_{n \rightarrow \infty} P_n(\hat{F}_n)$ بر فضای توابع F تعریف می شود و \hat{F}_n تابع توزیع تجربی است. بر اساس نظریه مذکور، می توان تابع توزیع توأم X_i را از طریق شرطی نمودن بر توزیع F و سپس انتگرال گیری تجزیه کرد. شایان ذکر است که Q تابع توزیع پیشین در نظر گرفته می شود.

تعریف ۳ (نظریه بیزی): طبق تعریف احتمال شرطی داریم: $p(B_i | A) = \frac{p(B_i A)}{p(A)}$ همچنین، می توان احتمال A را به صورت زیر تعریف کرد:

$$p(A) = \sum_i p(AB_i) = \sum_i p(A | B_i) p(B_i) \quad \text{رابطه ۴}$$

بنابراین، با توجه به موارد فوق، می توان نوشت:

$$p(B_i | A) = \frac{p(B_i) p(A | B_i)}{\sum_j p(B_j) p(A | B_j)} \quad \text{رابطه ۵}$$

خوشه بندی سلسله مراتبی بیزی

تحلیل چندمتغیری بیزی، یکی از فنون مهم در آمار به شمار می آید که می تواند با لحاظ توزیع پیشین برای داده ها بر دقت مدل و محاسبات تأثیرگذار باشد (لین، ۲۰۱۳). رویکرد بیزی در خوشه بندی سلسله مراتبی، با در نظر گرفتن مقدار احتمال تابع توزیع پسین، برای داده های نزدیک به مرز خوشه ها، حائز اهمیت است. علاوه بر این، رویکرد بیزی در خوشه بندی سلسله مراتبی به جای استفاده از فواصل اقلیدسی - که گاهی به دلیل نوع داده ها (داده های کیفی یا آمیخته) محدودیت دارد - معیاری بر اساس مقدار احتمال، برای تصمیم گیری در خصوص ادغام خوشه ها ارائه می دهد که از جمله مزایای آن محسوب می شود (هلر و قهرمانی، ۲۰۰۵).

همان‌طور که رابطه ۱ رویکرد بیزی در خوشه‌بندی بر مبنای مدل را به روشنی نشان می‌دهد، در این تحقیق، از رویکرد بیزی برای خوشه‌بندی استفاده شده و نتیجه محاسبات با خوشه‌بندی سلسله‌مراتبی کلاسیک مقایسه شده است.

با توجه به پیچیدگی توزیع پسین در خوشه‌بندی بیزی، می‌توان از تابع چگالی پسین حاشیه‌ای در خوشه‌بندی تراکمی، به‌عنوان معیاری برای تشابه/فاصله استفاده کرد و درخت سلسله‌مراتبی دندوگرام^۱ را ایجاد نمود. توزیع پسین حاشیه‌ای که معمولاً از طریق روش زنجیره مارکوف مونت کارلو محاسبه می‌شود، معیار ادغام و نقطه قطع دندوگرام را تعیین می‌کند (هلر و قهرمانی، ۲۰۰۵). در این روش، مقدار احتمال توزیع پسین در یک نقطه افزایش می‌یابد تا خوشه‌ها ادغام شوند، سپس مقدار آن کاهش می‌یابد. به این ترتیب بهترین دسته‌بندی خوشه‌ای ایجاد می‌شود (پرتوی‌نیا، ۲۰۰۹).

جزئیات الگوریتم متروپلیس هستینگ^۲ که یکی از روش‌های پرکاربرد در زنجیره مارکوف مونت کارلو (MCMC) است، در تحقیقات ژلمن، کارلین، استرن و روبین (۱۹۹۵) و نیز گیلکز، ریچاردسون و اشپگل هالتر (۱۹۹۶) استفاده شده است.

در ادامه، تابع توزیع پیشین برای دسته‌بندی داده‌ها و نیز توزیع پسین در خوشه‌بندی سلسله‌مراتبی بیزی ارائه می‌شود.

توزیع پیشین در خوشه‌بندی سلسله‌مراتبی بیزی

با فرض تابع توزیع پیشین تعویض‌پذیر، کافی است که برای T_c توزیع پیشین در نظر گرفته شود که T_c تعداد مشاهدات در خوشه c است. بنابراین داریم:

$$f(d) = \Pr(T_1, \dots, T_c | C) \Pr(C) \quad \text{رابطه ۶}$$

d برچسب داده‌ها در خوشه مربوطه است؛ به‌صورتی که داده‌های یک خوشه برچسب یکسانی دارند. با در نظر گرفتن تابع توزیع پیشین از نوع یکنواخت گسسته برای تعداد کل خوشه‌ها، به‌صورت $\Pr(C = c) = \frac{1}{T}$, $c = 1, \dots, T$ و تابع توزیع دیریکله چندجمله‌ای یکنواخت برای تعداد کل مشاهدات به شرط تعداد کل خوشه‌ها، خواهیم داشت (هرد، هولمز و استفانز، ۲۰۰۶).

1. Dendrogram
2. The Metropolis – Hasting Algorithm

$$f(d) = \Pr(T_1, \dots, T_c, C) \propto \frac{(C-1)! T_1! \dots T_c!}{(T+C-1)!} \quad \text{رابطه ۷}$$

توزیع پسین در خوشه‌بندی سلسله‌مراتبی بیزی

متغیر تصادفی y_{vct} نشان‌دهنده داده‌های خوشه‌بندی شده t ($t = 1, \dots, T_c$) در c خوشه ($c = 1, \dots, C$) برای متغیر پیوسته v ($v = 1, \dots, V$) است. توزیع پسین حاشیه‌ای در خوشه‌بندی را می‌توان به صورت زیر نوشت:

$$f(d | y) = k^{-1} f(y | d) f(d) \quad \text{رابطه ۸}$$

به دلیل آن که $K > 0$ برای تعداد مشخص، عددی ثابت است، می‌توان از آن چشم‌پوشی کرد. بنابراین، برای محاسبه مقدار تابع توزیع پسین حاشیه‌ای، تنها کافی است تابع توزیع پیشین و تابع توزیع حاشیه‌ای محاسبه شوند. تابع توزیع پیشین در رابطه ۶ تعریف شد و آن را به صورت حاصل ضرب تابع توزیع یکنواخت گسسته در تابع توزیع دیریکله چندجمله‌ای یکنواخت در نظر می‌گیریم. تابع چگالی توأم داده‌ها با خوشه به صورت زیر به دست می‌آید:

$$f(y | d) = \prod_{c=1}^C f(y_c) = \prod_{v=1}^V \prod_{c=1}^C f(y_{vc}) \quad \text{رابطه ۹}$$

باتوجه به موارد یاد شده، به طور خلاصه می‌توان گفت، به منظور انجام خوشه‌بندی سلسله‌مراتبی بیزی، ابتدا هر داده را به عنوان یک خوشه در نظر می‌گیریم، به همین دلیل تعداد خوشه‌ها برابر تعداد کل داده‌ها خواهد بود ($C = T$). همچنین، تعداد داده‌ها در خوشه c به صورت $T_c = 1$ برای تمام خوشه‌ها $c = 1, \dots, C$ است. در گام بعدی، ادغام دوبه‌دوی خوشه‌ها انجام می‌شود؛ به طوری که برای هر ادغام دوتایی، مقدار تابع توزیع پسین بر اساس رابطه ۸ محاسبه شده و ادغامی که باعث ایجاد بیشترین مقدار در رابطه ۸ شود، تعیین می‌شود. همچنین، برای تعیین ارتفاع درخت دندوگرام از $g_c = \log f(d | y)$ استفاده می‌شود. اگر بهترین ادغام بر اساس مقدار (رابطه ۸)، تعیین دو خوشه c_1 و c_2 برای ایجاد خوشه جدید باشد، خواهیم داشت: $T_c = T_{c_1} + T_{c_2}$. این فرایند مجدد ادامه می‌یابد تا تمام خوشه‌ها ادغام شده و تمام داده‌ها در خوشه‌های متناسب قرار گیرند. واضح است که برای ادغام خوشه‌ها، g_c باید دارای بیشترین مقدار باشد.

شبکه عصبی

شبکه عصبی، از رایج‌ترین تکنیک‌های داده‌کاوی برای برآورد و پیش‌بینی به‌شمار می‌رود. از ساختارهای رایج و معروف که در برآورد و پیش‌بینی، کاربرد بسیاری داشته و در این تحقیق نیز استفاده می‌شود، شبکه چندلایه پرسپترون است. برای ساخت یک مدل شبکه عصبی و استفاده از آن، باید تعداد لایه‌ها و نورون‌های هر لایه در شبکه، نوع شبکه و توابع فعال‌کننده مشخص گردند تا پس از آن مقدار وزن‌های شبکه برای نمونه‌های متعدد طی فرایند آموزش، اصلاح شوند و شبکه از این طریق فرایند یادگیری را طی کند تا برای داده‌های ورودی جدید، عملکرد خوبی داشته باشد.

یافته‌های پژوهش

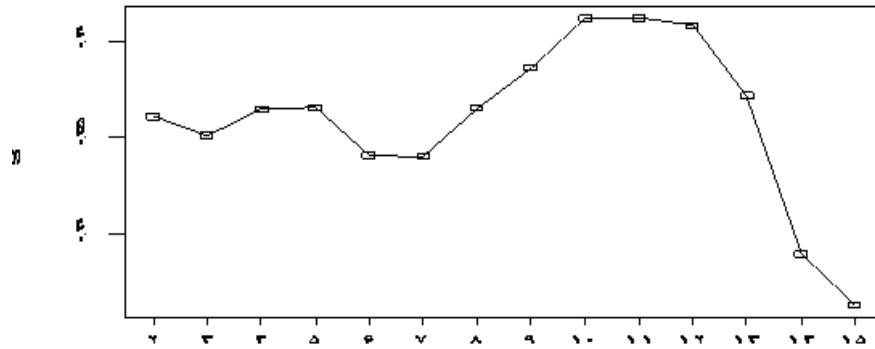
نتایج خوشه‌بندی شعب به روش سلسله‌مراتبی

با توجه به نوع داده‌های مورد استفاده در خوشه‌بندی شعبه‌ها، با استفاده از ضریب تشابه گوور^۱ (گوور، ۱۹۷۱) و انجام خوشه‌بندی سلسله‌مراتبی، شعبه‌ها به خوشه‌های متشابه دسته‌بندی شدند. اعتبار خوشه‌ها با استفاده از شاخص سیلهوتت روسو^۲ (کافمن و روسو، ۱۹۹۰) اندازه‌گیری شده است. افزایش مقدار شاخص سیلهوتت روسو، بیان‌کننده تعداد خوشه‌های بهینه‌تر است. جدول ۱ و شکل ۱، مقدار شاخص را برای تعداد خوشه‌های متفاوت نشان می‌دهند. همان‌طور که شکل ۱ نشان می‌دهد، بیشترین مقدار شاخص سیلهوتت، برای تعداد ۱۱ خوشه است، از این رو عملکرد خوشه‌بندی برای تعداد ۱۱ خوشه، نسبت به سایر تعداد خوشه‌ها بهینه است. با استفاده از الگوریتم دورترین همسایه در روش خوشه‌بندی سلسله‌مراتبی با تعداد ۱۱ خوشه، شماره شعبه‌ها در داده‌ها استخراج شدند که در شکل ۲ مشاهده می‌شود.

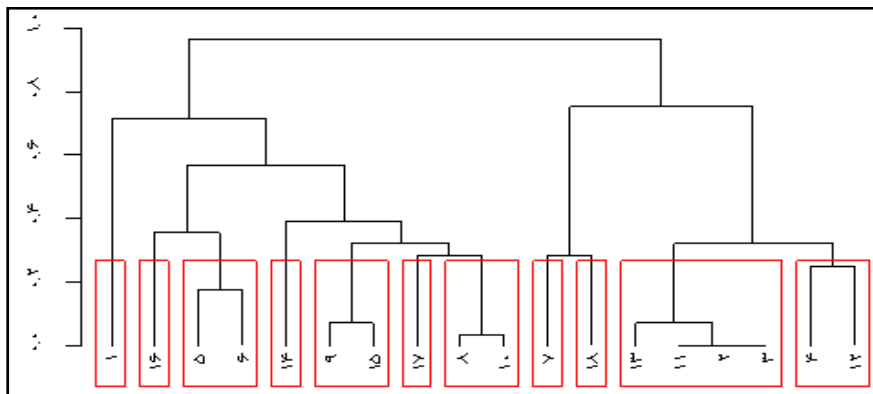
جدول ۱. مقادیر شاخص‌های ارزیابی اعتبار خوشه‌ها با تعداد متفاوت

تعداد خوشه‌ها (K)	۲	۳	۴	۵	۶	۷	۸
مقدار شاخص سیلهوتت روسو	۰/۳۶۱	۰/۳۵۱	۰/۳۶۴	۰/۳۶۵	۰/۳۴۰	۰/۳۴۰	۰/۳۶۵
تعداد خوشه‌ها (K)	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
مقدار شاخص سیلهوتت روسو	۰/۳۸۶	۰/۴۱۱	۰/۴۱۲	۰/۴۰۸	۰/۳۷۱	۰/۲۸۹	۰/۲۶۲

1. Gower coefficient of similarity
2. Rousseeuw's Silhouette index



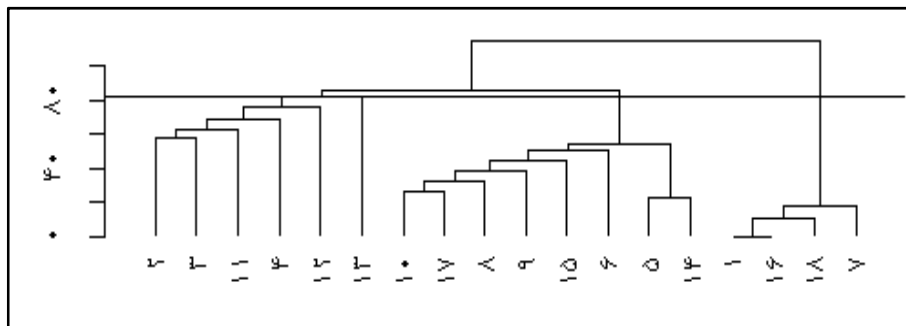
شکل ۱. نمودار شاخص سیلوهنت برای تعداد خوشه‌های مختلف



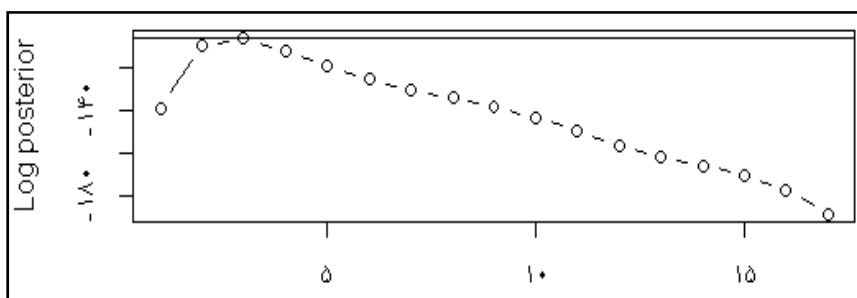
شکل ۲. نمودار خوشه‌بندی سلسله‌مراتبی

نتایج خوشه‌بندی شعبه‌ها به روش بیزی

در خوشه‌بندی شعبه‌ها به روش بیزی، داده‌ها از طریق تقسیم هر داده در هر متغیر به میانگین داده‌های موجود در آن متغیر، استاندارد شده و به داده‌های پیوسته تبدیل شدند. سپس با روش‌های ارائه شده در بخش‌های قبل، خوشه‌بندی شدند. خوشه‌بندی داده‌های مربوط به ۱۸ شعبه بانک تجارت، با استفاده از روش خوشه‌بندی سلسله‌مراتبی بیزی در شکل ۳ نشان داده شده است. همچنین براساس شکل ۴، سه خوشه دارای بیشترین مقدار لگاریتم تابع توزیع احتمال پسین است که تعداد خوشه بهینه در این روش را نشان می‌دهد.

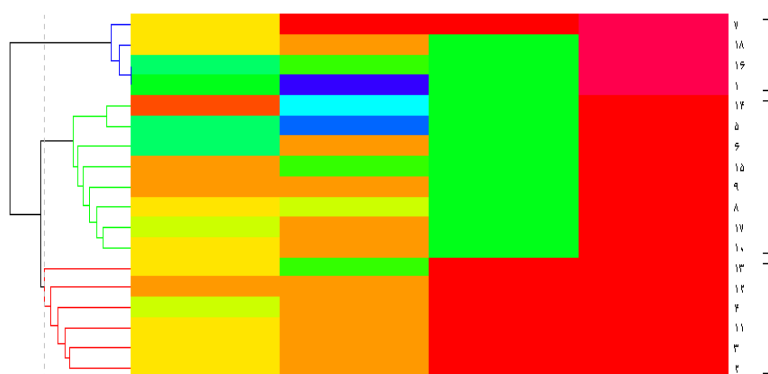


شکل ۳. نمودار خوشه‌بندی سلسله‌مراتبی بیزی



شکل ۴. نمودار مقدار لگاریتم توزیع پسین داده‌ها براساس تعداد خوشه‌های مختلف

شکل ۵ نیز، شماره برچسب داده‌های شعب بانک را در هر یک از سه خوشه نشان می‌دهد.



شکل ۵. نمودار برچسب داده‌های شعب بانک در هر یک از سه خوشه

در جدول ۲، شماره خوشه شعب بانک تجارت براساس دو روش خوشه‌بندی درج شده است.

جدول ۲. شماره خوشه شعب بانک تجارت براساس دو روش خوشه‌بندی

کد شعبه	شماره خوشه براساس روش سلسله‌مراتبی	شماره خوشه براساس روش سلسله‌مراتبی بیزی
۴۵۰	۱	۱
۶۶	۲	۳
۱۰۷	۲	۳
۱۳۵	۳	۳
۱۶۵	۴	۲
۵۴	۴	۲
۳۱۷	۵	۱
۴۱۱	۶	۲
۲۱	۷	۲
۳۱۲	۶	۲
۳۶۲	۲	۳
۳۹۲	۳	۳
۷۰۶	۲	۳
۱۰۱۹	۸	۲
۰	۷	۲
۵۶	۹	۱
۲۳۱	۱۰	۲
۲۷۰	۱۱	۱

نتایج به‌دست آمده از شبکه عصبی براساس نتایج خوشه‌بندی

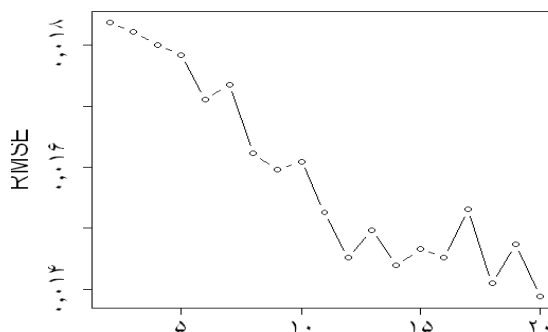
با مدنظر قراردادن نتایج به‌دست آمده از خوشه‌بندی شعب، در گام بعد با استفاده از شبکه عصبی به برآورد وجه نقد آورده توسط مشتری به شعبه و نیز وجه نقد مصرفی شعبه پرداخته می‌شود. به‌منظور افزایش دقت و سرعت در آموزش شبکه عصبی، ابتدا تمام داده‌ها با استفاده از رابطه ۱۰ استاندارد شدند، به‌طوری که تمام داده‌ها بین صفر و یک قرار گرفتند.

$$z_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad \text{رابطه ۱۰}$$

در این رابطه، z_i داده استاندارد شده؛ x_i داده استفاده شده و x_{min} و x_{max} به‌ترتیب بیشترین و کمترین داده در هریک از متغیرهاست. ۷۰ درصد داده‌ها به‌منظور آموزش شبکه و ۳۰ درصد

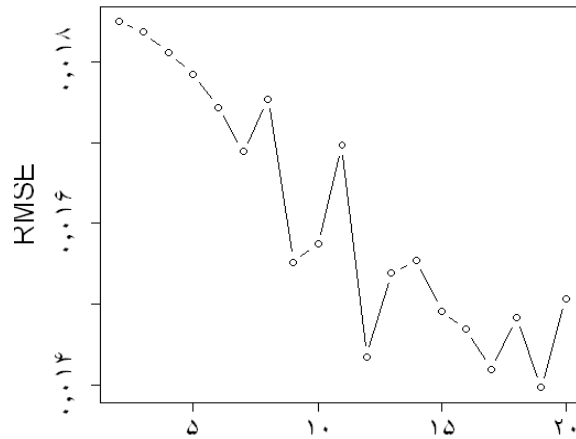
داده‌ها نیز به‌منظور آزمون شبکه به‌طور تصادفی استخراج و استفاده شدند. لایه ورودی شبکه عصبی برای هریک از برآوردها (برآورد وجه نقد مصرفی شعبه و برآورد وجه نقد آورده مشتری) دارای ۵ نورون شامل متغیرهای «روز کاری»، «روز هفته»، «روز پرداخت حقوق/ واریز یارانه/ واریز سود»، «روز تعطیل و مناسبت‌های رسمی» و «شماره خوشه» است. لایه خروجی نیز برای هریک از برآوردها، دارای یک نورون شامل متغیرهای «وجه نقد مصرفی شعبه» و «وجه نقد ورودی به شعبه» است. متغیرهای روز کاری، روز هفته، روز پرداخت حقوق/ واریز یارانه/ واریز سود، روزهای تعطیل و مناسبت‌های رسمی، متغیرهای تقویمی هستند، به طوری که متغیر «روز کاری»، براساس ترتیب روزهای کاری از ابتدای ماه تا آخرین روز کاری هر ماه و متغیر «روز هفته»، براساس ترتیب روزهای هفته کدگذاری شده‌اند. همچنین، متغیر «روز پرداخت حقوق/ واریز یارانه/ واریز سود»، در روزهای اول تا سوم و نیز بیستم تا بیست و پنجم همراه وزن‌دهی شدند. روزهای کاری قبل و پس از تعطیلات رسمی نیز به‌دلیل تقاضای بیشتر وجه نقد در متغیر «روز تعطیل و مناسبت‌های رسمی»، مدنظر قرار گرفته‌اند. همچنین متغیر «شماره خوشه» براساس نتایج بخش مربوطه استفاده شده است. توابع فعال‌سازی نیز در لایه پنهان، تانژانت هیپربولیک و در لایه خروجی، خطی در نظر گرفته شده است.

الف) خروجی شبکه عصبی برای برآورد وجه نقد ورودی به شعبه با استفاده از نتایج خوشه‌بندی سلسله‌مراتبی: از آنجا که برای تعیین تعداد نورون‌ها در لایه پنهان از روش آزمون و خطا استفاده می‌شود، در این تحقیق با رویکرد پیشرو، ابتدا ۲ نورون در لایه پنهان را در نظر گرفته و با افزایش تعداد نورون‌ها و محاسبه معیار خطا، شامل ریشه میانگین توان دوم خطا (RMSE)، در نهایت ۲۰ نورون در لایه پنهان که به ایجاد شبکه‌ای با کمترین مقدار خطا منجر شده است، انتخاب شد (شکل ۶).



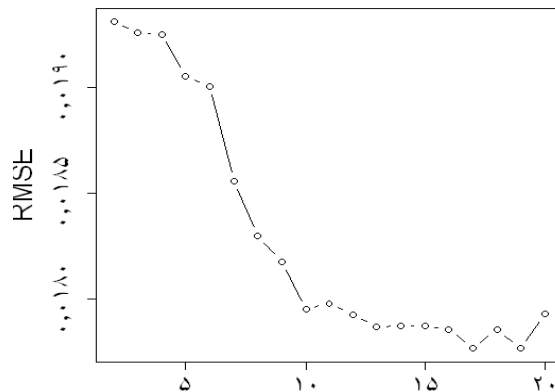
شکل ۶. نمودار مقدار معیار خطا نسبت به تعداد نورون‌ها در لایه پنهان برای برآورد وجه نقد آورده شده به شعبه (با نتایج خوشه‌بندی معمولی)

ب) خروجی شبکه عصبی برای برآورد مصرف وجه نقد با استفاده از نتایج خوشه‌بندی سلسله‌مراتبی: در این خصوص، ۱۹ نورون در لایه پنهان که به ایجاد شبکه‌ای با کمترین مقدار خطا منجر شده، انتخاب شده است (شکل ۷).



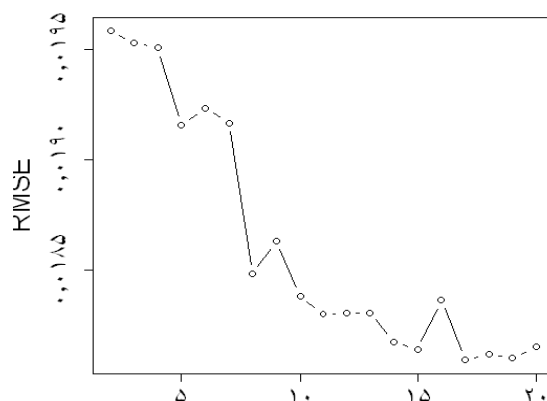
شکل ۷. نمودار مقدار معیار خطا نسبت به تعداد نورون‌ها در لایه پنهان برای برآورد مصرف (با نتایج خوشه‌بندی معمولی)

ج) خروجی شبکه عصبی برای برآورد آورده وجه نقد به شعبه با استفاده از نتایج خوشه‌بندی سلسله‌مراتبی بیزی: ۱۹ نورون در لایه پنهان که به ایجاد شبکه‌ای با کمترین مقدار خطا منجر شده، انتخاب شده است (شکل ۸).



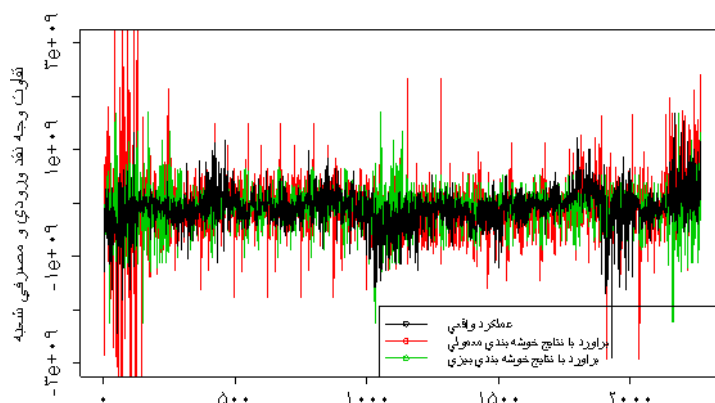
شکل ۸. نمودار مقدار معیار خطا نسبت به تعداد نورون‌ها در لایه پنهان برای برآورد وجه نقد ورودی به شعبه (با نتایج خوشه‌بندی بیزی)

د) خروجی شبکه عصبی برای برآورد مصرف وجه نقد با استفاده از نتایج خوشه‌بندی سلسله‌مراتبی بیزی: ۱۷ نورون در لایه پنهان که به ایجاد شبکه‌ای با کمترین مقدار خطا منجر شده، انتخاب شده است (شکل ۹).



شکل ۹. نمودار مقدار معیار خطا نسبت به تعداد نورون‌ها در لایه پنهان برای برآورد مصرف (با نتایج خوشه‌بندی بیزی)

به دلیل حجم زیاد داده‌ها برای ۱۸ شعبه بانک در ۵ ماه، از ارائه تمام مقادیر خودداری شده است. از این رو شکل ۱۰، مقدار واقعی تفاوت وجه نقد ورودی و خروجی (وجه نقد مورد نیاز) و نیز تفاوت مقادیر برآورد شده (وجه نقد مورد نیاز برآورد شده) شعبه‌ها را طی ۵ ماه نشان می‌دهد.



شکل ۱۰. نمودار عملکرد واقعی شعب طی ۵ ماه در مقایسه با نتایج برآوردهای انجام شده از طریق شبکه عصبی کلاسیک با لحاظ نتایج خوشه‌بندی معمولی و خوشه‌بندی بیزی

بر اساس جدول ۳، میانگین قدرمطلق خطای برآوردهای انجام شده برای تمام ۱۸ شعبه نمونه آماری در بازه زمانی ۵ ماه (از اردیبهشت تا شهریور ۱۳۹۳) برای نتایج حاصل از شبکه عصبی و براساس دو روش خوشه‌بندی معمولی و بیزی، به ترتیب برابر ۰/۰۰۲۶۷ و ۰/۰۰۲۵۹ به دست آمد. این مقادیر نشان می‌دهند، برآوردهای انجام شده با استفاده از شبکه عصبی و با لحاظ نتایج خوشه‌بندی شعب با رویکرد بیزی، نسبت به نتایج خوشه‌بندی شعب با روش معمول دقت بیشتری دارد.

جدول ۳. مقادیر خطا برای تفاوت وجه نقد ورودی و خروجی برآورد شده در ۱۸ شعبه بانک در ۵ ماه

۰/۰۰۲۵۹	مقدار میانگین قدرمطلق خطا برای تفاوت وجه نقد ورودی و خروجی برآورد شده شعب براساس نتایج خوشه‌بندی بیزی
۰/۰۰۲۶۷	مقدار میانگین قدرمطلق خطا در تفاوت وجه نقد ورودی و خروجی برآورد شده شعب براساس نتایج خوشه‌بندی معمولی

نتیجه‌گیری

در این پژوهش ۱۸ شعبه بانک تجارت در بازه زمانی ۵ ماه، با توجه به وجود تنوع بین آنها، به کمک دو روش خوشه‌بندی سلسله‌مراتبی و خوشه‌بندی سلسله‌مراتبی با رویکرد بیزی، در خوشه‌های متشابه دسته‌بندی شدند؛ سپس با مد نظر قرار دادن نتایج هر دو روش خوشه‌بندی و استفاده از معیار خطا، ساختار شبکه عصبی برای هریک از مقادیر وجه نقد مصرفی در شعبه و نیز وجه نقد آورده شده توسط مشتری به شعبه، تعیین شد تا از طریق آنها وجه نقد لازم برای شعبه‌ها محاسبه شود. براساس نتایج، میانگین قدرمطلق خطای برآوردهای انجام شده برای تمام ۱۸ شعبه بانک تجارت (نمونه آماری) در بازه زمانی ۵ ماه با استفاده از شبکه عصبی و براساس دو روش خوشه‌بندی معمولی کلاسیک و بیزی، به ترتیب ۰/۰۰۲۶۷ و ۰/۰۰۲۵۹ به دست آمد که نشان می‌دهد برآوردهای انجام شده با استفاده از شبکه عصبی و لحاظ نتایج خوشه‌بندی شعب با رویکرد بیزی، دقت بیشتری نسبت به نتایج خوشه‌بندی شعب با روش معمول دارد.

پیشنهادها

۱. استفاده از روش شبکه‌های عصبی می‌تواند به مشکل بیش‌برازش بینجامد، به طوری که آموزش شبکه با حداقل خطا انجام شود، ولی شبکه در داده‌های آزمون و پیش‌بینی داده‌های جدید عملکرد خوبی نداشته باشد. در این خصوص، به کارگیری رویکرد بیزی در شبکه‌های

عصبی (علاوه بر خوشه‌بندی بیزی) می‌تواند مشکل بیش‌برازش را حل کند و حتی ممکن است بر دقت مدل نیز تأثیرگذار باشد که این موضوع به‌عنوان تحقیقات آتی پیشنهاد می‌شود.

۲. با توجه به آن که یکی از مسائل مهم برای مدیران و رؤسای تمام بانک‌ها، پیش‌بینی مقدار وجه نقد دریافتی / پرداختی از / به خزانه است تا هزینه نقل و انتقال پول به حداقل برسد، تعیین وجه نقد دریافتی / پرداختی از / به خزانه توسط شعب نیز حائز اهمیت است و می‌تواند در مطالعات آتی مدنظر قرار گیرد.

فهرست منابع

بهشتی، ش. (۱۳۸۹). یک مدل برای پیش‌بینی الزامات نقدینگی شعب بانک و پیاده‌سازی آزمایشگاهی با استفاده از شبکه‌های عصبی. پایان‌نامه کارشناسی ارشد، تهران، دانشکده مدیریت و حسابداری. دانشگاه علامه طباطبایی.

Beheshti, Sh. (2010). *A model to predict the liquidity requirements of bank branches and implemented experimentally using Neural Networks*. MSc thesis, Faculty of Management and Accounting, Allameh Tabataba'i University. (in Persian)

Bernardo, J.M & Giron, J. (1988). A Bayesain Approach to cluster Analysis, *Questiio*, 12(1), 97-112.

Bramer, M. (2007). *Principles of Data Mining*. Springer Verlag Berlin Heidelberg.

Blomstedt, P., Tang, J., Xiong, J., Granlund, Ch. & Corander, J. (2015). A Bayesian Predictive Model for Clustering Data of Mixed Discrete and Continuous Type. *IEEE transactions on Pattern analysis and machine intelligence*, 37(3), 489-498.

Bouveyron, C. & Brunet-Saumard, C. (2014). Model-based clustering of highdimensional data: A review. *Computational Statistics and Data Analysis*, 71, 52-78.

Cardona, L. & Amaya, M. L. (2012). Cash management cost reduction using data mining to forecast cash demand and LP to optimize resources. *Mathematic. Comp.*, 4 (2), 127-134.

Franzen, J. (2008). *Bayesain cluster Analysis*. Doctoral dissertation. Department of statistics, Stockholm University.

- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Gower, J. (1971). A general coefficient of similarity and some of its properties. *Bio Metrics*, 27(4), 857-874.
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufman.
- Heard, A.N. , Holmes, Ch.C. & Stephens, A. (2006). A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheline Mosquitoes: An Application of Bayesian Hierarchical Clustering of Curves. *American Statistics Association*, 101(473), 18-29.
- Heller, K. A. & Ghahramani, Z. (2005). Bayesian hierarchical clustering. *In Twenty-second International Conference on Machine Learning*.
- Kauffman, L. & Rousseeuw, P. (1990). *Finding groups in Data: an introduction to cluster analysis*, New York: Wiley.
- Lin, M.Y. (2013). *Bayesian Statistics*. Technical Report No. 2, Boston University, Department of health and Management.
- Noorbakhsh, I., Heydari, H. & Zavarian, Z. (2010). A diffusion approximation model for vault cash management. *7th Report, Monetary and Banking Research Institute*, IRI Central Bank. (in Persian)
- Partovi-Nia, V. (2009). *Fast High-Dimensional Bayesian Classification and Clustering*. Ph.D. thesis, Ecole Polytechnique Fédérale de Lausanne.
- Partovi-Nia, V. & Davison, A. C. (2012). High-Dimensional Bayesian Clustering with Variable Selection: The R Package bclust. *Journal of Statistical Software*, 47(5), 1-22.
- Premchand, K. & Walia, E. (2006). Cash Forecasting: An Application of Artificial Neural Networks in Finance, *International Journal of Computer Science & Applications*, 3 (1), 61-77.
- Sandipan, R. (2010). Determining Optimal Cash Allocation at ICICI Bank Branches. *Mumbai: ICICI Bank, SAS Global Forum*.

برآورد وجه نقد ورودی و خروجی شعب بانک تجارت برای محاسبه... ۶۰

Wang, P. (2008). *Clustering and Classification Techniques for Nominal Data Application*. Dissertation, Department of Electronic and Engineering, City University of Hong Kong.